大数据出版在童书选题策划中的应用

许　静
（河南科学技术出版社有限公司，河南 郑州 450016）

摘　要：伴随着新媒体技术高速发展，使得在当下的大数据出版发展中，与出版业务实现了深入的融合。因此，这样的技术发展背景下，传统的出版模式已经无法充分满足现代图书出版的发展需求。在本文分析中，主要针对大数据出版技术在童书选题策划当中的具体应用进行详细的分析和阐述，并基于一些实际的选题策划实例为例，明确在大数据出版模式中对于童书选题策划的重要价值。

关键词：新媒体技术；大数据出版；童书；数字化阅读资源；信息化时代　　**中图分类号：**G232.1　　**文献标识码：**A
文章编号：1671-0134（2021）07-111-03　　**DOI：**10.19483/j.cnki.11-4653/n.2021.07.033
本文著录格式：许静．大数据出版在童书选题策划中的应用 [J]. 中国传媒科技，2021（07）：111-113.

chinaXiv:202310.00979v1

大数据是一种在信息技术的发展背景下，所形成的一种全新的思维体系，近些年来，这一技术已经受到了出版领域的格外重视。同方知网（北京）技术有限公司提出了“大数据”的出版理念，但是到目前始终没有形成一个较为完整的理论体系。在我国，童书出版领域市场较大，并且这一领域已经进入到了出版业发展的重要阶段，因此仅仅遵循着传统的选题策划具有局限性，会严重地影响到未来市场规模的扩展。

1. 大数据出版

对大数据出版工作，是一种基于传统的出版模式，与数字出版工作实现结合的新型出版模式。在这样的出版模式下，可以有效扩展传统出版资源的利用率，以此进一步完善数字化阅读模式，在数字化阅读方式中，对其阅读行为实现全面的整合分析，以此有效实现数据资源的共享和整合。[1] 另外，还可以充分将出版行业的发展，进行线上与线下的深度融合，进而有效解决出版、发行以及销售当中存在的断裂问题。在这样的发展中，便可以很好地利用先进的技术方式，实现数据化方面的全面升级。这样的出版方式下，是一种对传统出版模式的全面进步和革新，有着里程碑式的进步。大数据出版模式的应用中，可以很好地实现信息的高效率存储、传播以及创新，不仅仅对当下的信息实现全面的整合，同样也适用于过去的信息数据处理工作当中。在没有研发出大数据出版模式之前，所使用的传统出版，或者开展数字出版的过程中，由于每一个环节都缺乏相应的集成性和预算筹资方面的能力，就会导致大量资源和信息出现浪费和丢失的问题。在当下大数据技术的广泛应用下，使得不再出现信息孤岛的问题，可以很好地将不同的信息数据连接起来，形成大数据出版的模式。[2]

2. 出版企业利用大数据的具体方式

在对大数据长期研究中，一些研究机构对大数据进行了全面的定义，首先大数据是一种需要进行处理之后，才具备着较高的洞察力、发现力以及流程优化的能力，并且呈现出高增长率、多样化的信息资产特征。由于现今社会进入到信息化的时代当中，使得数据规模呈现出高速增长态势，需要处理的数据资源越来越多，这对传统的信息数据管理体系造成了较大的处理压力。传统的信息管理模式越发地凸显出处理模式的局限性，因此就需要实现“大数据”的处理模式。在当下社会中形成的大数据，主要体现出信息数据大、类型多、时效性以及价值高的特征。因此，在这样的出版企业的发展中，可以很好地将产业的未来发展与大数据技术进行深入的融合，以此推动产业的发展。[3]

2.1　Web 文本挖掘

大数据的核心往往有着较高的数据价值。社会的发展与信息技术有着越来越紧密地联系，以此形成了大量的信息数据，并出现了大量半结构化，或者一些非结构化的文本信息。在各个行业的发展中，更加需要研发出便捷的工具，有效地将大量的数据信息，转变成人们可以理解的知识内容。Web 文本挖掘技术的出现，正是为了解决这样的信息价值需求，所形成的大数据技术。[4]

文本挖掘技术的发展，也是当下数据挖掘技术当中十分重要的一种技术分支。在实际的应用中，主要是对文本数据库当中的知识内容进行挖掘，有效地让大量的文本实现集合，并在语料库当中，对一些可以分析的、并具有较高价值性的信息，实现深入地挖掘。在互联网技术高速发展背景下，让 Web 应用大面积的普及到了出版行业的发展当中，但是一定程度上也让信息数据出现了更多的类型。在未来的发展中，就更加需要对 Web 进行深度的文本挖掘，以此在海量的数据信息中，提取出更多有价值的信息数据信息。[5]

2.2　利用 Web 文本数据

在网络中有着大量的数据信息，甚至一些畅销书在出版的过程中，涉及到大量的评论信息。为此，对这种

ChinaXiv合作期刊

chinaXiv:202310.00979v1

Web 文本挖掘技术的应用，可以有效地对出版物进行评论方面的集成化处理，以此提取出有价值性的信息。在实际的处理中，主要是利用对图书方面开展在线评论，或者实现管理工作的挖掘，让童书出版单位，可以对在控制成本的前提下，顺利地完成对各种类型童书评论信息的整理和分析。另一方面，也使得出版单位时刻掌握用户对童书的想法，这样在选题策划的过程中，便专注市场的需求，从迎合市场需求的角度出发，选择适合用户阅读的题材。另外，这样的出版模式，也有效地让出版单位实现了各种资源的全面整合，在获取到大量知识资源的时候，也培养了持续性的竞争优势。

而对企业而言，在一些童书的销售网站的建立中，也可以很好地利用互联网销售的平台，实现对消费者知识管理体系的构建，对不同客户的评论和意见实现集中化的整合分析，将消费者的评论逐渐转变成企业的知识，让出版机构在未来的童书出版中，可以基于用户的实际需求，做出合理的选题、出版方面的决策。而企业同时也可以对客户实现知识管理工作，针对用户的一些问题，提出针对性地解决意见。另外，在构建的客户知识管理系统中，也有着较强的用户需求同步的反应能力，更加有效的基于用户的实际需求，提升童书策划以及印制的能力。只有有效的提升出版机构与用户之间的联系，才可以保障在童书出版中，有着较强的决策能力，推动童书市场的发展。[6]

3. 大数据出版在童书选题策划中的应用

3.1 信息收集与预处理

对于传统数据库中完全结构化的数据而言，在数据中最大特征，就是其信息数据呈现出半结构化的特征。而基于对评论数据信息的深入挖掘角度出发，形成的 Web 数据信息中，其网页中所包含着的信息有着较高的挖掘价值，以此其网页中的主题信息都是重要的主题内容体现。但是对一些与主题内容不相关的导航条、广告等信息，都是一些噪音信息，不需要进行挖掘。因此，网页方面的净化处理，就是对这些噪声内容的清除，保障在网页当中仅仅保留一些与主题内容相关的信息数据。在实际的评论挖掘当中，首先就要开展网页的全面净化处理，因此就可以让网页当中的音频、视频、图片中的一些分本文的信息内容实现全面的净化处理。另一方面，还要对网页中的文字评论内容，进行深度的评论分析。

同时在评论页面的下载过程中，还要可以使用聚焦爬虫技术的方式。所谓聚焦爬虫技术，就是一种可以很好地实现网页的自动下载功能。在使用中，基于设定的内容，实现网页端信息的自动抓取。另一方面，还要有选择地让万维网上的一些网页，以及相关的链接进行深入地分析，以此有效的挖掘出更多的信息数据。之后，还要使对 Web 页面的语义结构，进行详细的描述，并设计出一些符合计算机自动处理，以及描述文件方面的信息内容。这样就可以在未来的使用中，可以持续性的对 Web 信息进行挖掘。[7]

3.2 图书特征提取和情感分析

在当下的市场竞争中，童书出版行业越发地激烈起来，以此也就在童书的选题策划过程时，面临着越来越大的压力。在过去的童书出版过程中，始终无法有效地解决粗放式、分散化的出版问题。加上营销产业模式的落后，使得童书出版的过程中，逐渐呈现出低俗化、功利化的发展趋势，这并不利于童书产业的健康发展。为此，为了推动童书出版行业的可持续发展，就要对选题策划的过程中，利用其大数据出版的技术优势性，能够对选题的未来发展做出科学合理的规划，以此解决各种潜在的问题。在满足我国儿童对童书的基本需求的前提下，还要充分利用大数据中的各种资源，保障童书编辑在实际的选题中，可以提升对大数据出版的重视程度。

在当下的出版发展中，移动阅读已经呈现出高速发展的态势。例如微信、微博这些多媒体平台的出现，更加推动了整个社会的移动阅读发展。首先，在我国的童书选题中，需要结合起社会的实际发展情况，实现动态化的选题。[8] 例如，在编辑的选题中，要首先将图书线上与线下的传播和营销方式结合起来。之后，还要充分地利用各个社会平台，实现大数据方面的动态监测，这样就可以在未来的出版中，明确出选题的主要方向。接着，还要进一步提升图书在设计以及印刷方面的品质。以此让童书具有重要的价值。[9] 读者对于出版机构信任，是一种有效提升图书内容以及质量的重要途径 . 为此利用大数据出版技术，便可以很好地对当下一些畅销类图书，在出版以及设计中，所使用的各种装帧和内容选题上的标准进行分析，这样才可以在未来的出版中，更加明确出版的目标。最后，还要将选题的途径实现多样化的发展。在当下的互联网高度普及的社会环境中，要有效地让选题内容不仅仅局限于一些传统的图书内容，还要结合起音乐、视频等各种类型的多媒体形式，为童书的出版构建出多元化的平台。对这样的数据库建立，可以较为有效地将内容传播与出版形成高度的融合。[10]

在大数据出版技术的使用中，可以很好明确出儿童对童书的关注点，进而保障读者在图书评论当中，逐渐形成较为明显的特征属性。但是对于不同特征的信息挖掘，还要对其进行深入的分析和处理，在保障对读者的每一个人关注点都进行明确之后，便基于各种类型的特征反映的态度，实现情感倾向性的明确。在情感词的划分之后，可以很好地形成各种类型的情感特征词汇，并形成诸多的情感强度词汇。对形成的情感词汇进行深入地分析中，便能够形成对消费者对童书的情感取向，客户在实际的评论中，对客户的消费行为起到了决定性的作用，只有让儿童对童书的内容有兴趣，才能推动童书的销售。这样在未来的童书选题上，就可以基于用户对

chinaXiv:202310.00979v1

童书内容的兴趣点，开展针对性的选题，保障基于用户的角度上，进行童书的出版。

3.3 童书选题的国际化发展

伴随着社会的发展和进步，使得我国当下的童书出版市场已经呈现出国际化的发展趋势。因此，在这样国际化的发展模式下，不同于传统的买卖版权的贸易形式，而是呈现出多样化的出版模式。例如，在图书的合作出书模式当中，就已经涉及各种国家的文化。二十一世纪出版社，就将波兰的画家请到南昌来，以此合作出版了一系列的童书作品，并在市场中受到了儿童的欢迎。同时，也邀请了一些日本地区的作家，共同完成了大量童书的出版和设计。在这样国际化的童书选题设计中，是一种对童书选题方面的创新发展，在扩展了儿童的知识面的同时，也极大地丰富了童书的内容，因此受到了市场的追捧。为此，在未来的童书选题策划过程中，要将选题视野进行扩展，积极地从一些优秀的历史、民族文化中进行取材，以此起到丰富童书内容的作用。互联网的发展推动了全世界经济与文化之间的沟通和交流，以此让人们更加关注其他文化的内容。在童书的选题策划上，积极地选择一些国际上的文化和内容，也有效地起到吸引儿童兴趣的作用，大大提升童书在市场同类产品的竞争力，是一种推动我国童书产业发展的重要方式。

4. 未来童书选题策划发展前景

多媒体技术的发展，使得出版社要在童书的设计中，不仅仅将选题的内容涉及到本领域的产业，而且加强同其他媒体类型的融合发展。当下的产业在信息技术的发展趋势下，形成了多元化的技术发展趋势，为此在童书的选题策划过程中，就要实现三个不同方面的转变。首先，要在童书整体选题策划中，可以实现专业化的发展，不仅仅要实现内容的专业化发展，同时还要在具体策划中，可以保障策划理念以及策划精神方面的专业化发展。要让更多专业的出版社，用先进的大数据技术，实现各种资源的整合和分析，以此在童书的出版中，推出更多质量性较高的童书出版产品。其次，还要积极地实现童书出版方面的创新，积极地与一些外国的童书出版物取得联系，保障同国际市场的衔接。这样就可以在童书的出版设计上，形成全新的发展思路，以此让童书的内容选题上，有着更多的可能性，让资源共享可以很好地推动行业者的发展，在我国的童书出版领域逐渐形成社群的联合，提升信息共享的能力。在发现和挖掘出有价值的童书选题信息时，形成一个发展创新的合力，共同推动整个行业的发展。在未来企业的发展中，为了不断地扩大产品的销量，以及提升企业的经营规模，就要加强同各个领域的合作和交流，以此有效地在未来的发展中，加强企业同企业之间的关联，在出版社、读者以及编辑之间，形成生态互联网发展模式。这种模式的出现也意味着我国的童书出版事业，呈现出规模化、标准化的发展态势。国家现代化的建设不仅仅需要社会的基础设施方面的完善，同时也需要各个产业都能够跟上现代化的发展脚步，有效地利用信息化技术，推动产业的现代化改革和进步。

结语

综上所述，在当下的发展中，信息技术深深地影响了诸多产业的发展，对童书的出版产业而言，可以有效地利用大数据出版的技术方式，为童书的选题策划提供相应的合理性建议。只有在这样的技术加持下，才能更好地推动整个图书产业的发展和进步。媒

参考文献

[1] 吴子琳 . 基于大众传播理论的选题辅助决策系统应用研究 [D]. 北京：北京印刷学院，2020.

[2] 吴晓红，王丹菲 . 现阶段中国主题出版选题研究 [J]. 安徽农业大学学报（社会科学版），2020（3）：116−122.

[3] 张聪聪 . 出版企业主题出版考评指标研究 [D]. 北京：北京印刷学院，2018.

[4] 李建红 . 中国主题出版研究（2003—2016 年）[D]. 武汉：武汉大学，2017.

[5] 霍文婵 . 出版项目选题分析与评价研究 [D]. 北京：北京理工大学，2016.

[6] 朱驰风 . 基于数字出版的 ERP 选题系统设计与实现 [D]. 武汉：武汉理工大学，2015.

[7] 卢芸 . 危机传播视阈下的“5·12”图书选题策划研究 [D]. 成都：西南交通大学，2012.

[8] 吴文峰 . 中国体育图书出版现状及发展战略研究 [D]. 北京：北京体育大学，2007.

[9] 康超 . 河北出版业选题特色研究 [D]. 保定：河北大学，2006.

[10] 谢洋 . 出版社选题决策模糊综合评估的方法研究 [D]. 成都：电子科技大学，2005.

作者简介：许静（1984−），女，河南固始，硕士研究生，编辑，研究方向：科普类、少儿类图书编辑出版。

（责任编辑：张晓婧）